

Claims

[Claim 1]

A database building apparatus for managing structured documents, the database building apparatus comprising:

5 an input document analysis portion for assigning a unique document number to each structured document and analyzing its structure;

 an element name registration portion for assigning a unique element name ID to each element name appearing in the structured
10 document based on results of the analysis performed by the input document analysis portion and registering the element name in an element name dictionary;

 an ancestral path name registration portion for assigning a unique ancestral path name ID to each ancestral path name
15 appearing in the structured document based on the results of the analysis performed by the input document analysis portion and registering the ancestral path name in an ancestral path name dictionary; and

 an appearance information registration portion for
20 registering element appearance information including at least information about a document number at which an element of interest appears, character position, ancestral path name ID, and order of branches in element appearance information storage portion using an element name ID as a key based on the results
25 of the analysis performed by the input document analysis portion

and for registering ancestral path appearance information including at least information about the document number at which the element of interest appears, character position, element name ID, and order of branches in an ancestral path appearance information storage portion using the ancestral path name ID as a key.

[Claim 2]

The database building apparatus of claim 1, further including an attribute name registration portion for assigning a unique attribute name ID to each attribute name appearing in the structured document based on the results of the analysis performed by the input document analysis portion and registering the attribute name in an attribute name dictionary,

wherein the appearance information registration portion registers attribute appearance information including at least information about a document number at which an attribute of interest appears, character position, ancestral path name ID, element name ID, and order of branches in an attribute appearance information storage portion using the attribute name ID as a key based on the results of the analysis performed by the input document analysis portion.

[Claim 3]

The database building apparatus of claim 1, wherein the appearance information registration portion registers text appearance information including at least information about

appearing document number, character position, ancestral path name ID, element name ID, attribute name ID, and order of branches regarding partial character strings extracted from element entity text and attribute values in text appearance information storage portion using the extracted partial character strings as keys based on the results of the analysis performed by the input document analysis portion.

[Claim 4]

The database building apparatus of claim 1, wherein the element appearance information includes at least information about a document number at which an element of interest appears, character position, ancestral path name ID, order of branches, and order of empty elements, and wherein the ancestral path appearance information includes at least information about the document number at which the element of interest appears, character position, element name ID, order of branches, and order of empty elements.

[Claim 5]

The database building apparatus of claim 2, wherein the element appearance information includes at least information about the document number at which the element of interest appears, character position, ancestral path name ID, order of branches, and order of empty elements;

wherein the ancestral path appearance information includes at least information about the document number at which

the element of interest appears, character position, element name ID, order of branches, and order of empty elements; and

wherein the attribute appearance information includes at least information about the document number at which the attribute
5 of interest appears, character position, ancestral path name ID, element name ID, order of branches, and order of empty elements.

[Claim 6]

The database building apparatus of claim 3,
10 wherein the element appearance information includes at least information about the document number at which the element of interest appears, character position, ancestral path name ID, order of branches, and order of empty elements;

wherein the ancestral path appearance information
15 includes at least information about the document number at which the element of interest appears, character position, element name ID, order of branches, and order of empty elements; and

wherein the text appearance information includes at least information about appearing document number, character position,
20 ancestral path name ID, element name ID, attribute name ID, order of branches, and order of empty elements regarding partial character strings extracted from element entity text and attribute values.

[Claim 7]

25 The database building apparatus of claim 1, wherein the

ancestral path name registration portion assigns a unique ancestral path name ID to each partial ancestral path name obtained by dividing each ancestral path name appearing in the structured document into more than one partial ancestral path name and registers the partial ancestral path name in the ancestral path name dictionary.

[Claim 8]

The database building apparatus of claim 1, further including an appearance information grouping portion for grouping entries having common values of more than one information item other than document number and character position regarding entries of the element appearance information registered in the element appearance information storage portion using the same element name ID as a key and entries of the ancestral path appearance information registered in the ancestral path appearance information storage portion using the same ancestral path name ID as a key.

[Claim 9]

A database search apparatus for managing structured documents, the database search apparatus comprising:

an element name dictionary in which a unique element name ID has been registered for each element name appearing in each structured document;

an ancestral path name dictionary in which a unique ancestral path name ID has been registered for each ancestral

path name appearing in the structured document;

an element appearance information storage portion in which element appearance information has been stored using an element name ID as a key based on results of analysis of the structured document, the element appearance information including at least information about a document number at which an element of interest appears, character position, ancestral path name ID, and order of branches;

an ancestral path appearance information storage portion in which ancestral path appearance information has been stored using an ancestral path name ID as a key based on the results of the analysis of the structured document, the ancestral path appearance information including at least information about the document number at which the element of interest appears, character position, element name ID, and order of branches;

a search condition input portion for entering a search formula;

a search condition analysis portion for converting the input search formula into an internal condition formula by referring to the element name dictionary and the ancestral path name dictionary; and

an appearance information acquisition portion for finding plural search results from element appearance information from the element appearance information storage portion and from ancestral path appearance information from the ancestral path

appearance information storage portion according to the internal condition formula output by the search condition analysis portion.

[Claim 10]

5 The database search apparatus of claim 9, further including:

 an attribute name dictionary in which attribute name IDs and corresponding attribute names are recorded; and

 an attribute appearance information storage portion in
10 which attribute appearance information is stored using the attribute name IDs as keys, the attribute appearance information including at least information about a document number at which an attribute of interest appears, character position, ancestral path name ID, element name ID, and order of branches;

15 wherein the search condition analysis portion converts a search formula entered from the search condition input portion into internal condition formulas while referring to the element name dictionary and the ancestral path name dictionary; and

 wherein the appearance information acquisition portion
20 finds plural search results from element appearance information from the element appearance information storage portion, ancestral path appearance information from the ancestral path appearance information storage portion, and attribute appearance information from the attribute appearance
25 information storage portion according to the internal condition

formula output by the search condition analysis portion.

[Claim 11]

The database search apparatus of claim 9, further including a text appearance information storage portion in which text appearance information is stored using extracted partial character strings as keys regarding the partial character strings extracted from element entity text and attribute values, the text appearance information including at least information about appearing document number, character position, ancestral path name ID, element name ID, attribute name ID, and order of branches;

wherein the appearance information acquisition portion finds plural search results from element appearance information from the element appearance information storage portion, ancestral path appearance information from the ancestral path appearance information storage portion, and text appearance information from the text appearance information storage portion according to the internal condition formula output by the search condition analysis portion.

[Claim 12]

The database search apparatus of claim 9, wherein the appearance information acquisition portion compares the number of entries of a specified element name ID in the element appearance information storage portion and the number of entries of a specified ancestral path name ID in the ancestral path appearance information storage portion, refers to appearance information

having the fewer number of entries, and finds plural search results.

[Claim 13]

A method of constructing a database for managing structured documents, the method comprising the steps of:

assigning a unique document number to each structured document and analyzing its structure;

assigning a unique element name ID to each element name appearing in the structured document based on results of the analysis and registering the element name in an element name dictionary;

assigning a unique ancestral path name ID to each ancestral path name appearing in the structured document based on results of the analysis and registering the ancestral path name ID in an ancestral path name dictionary; and

registering element appearance information including at least information about a document number at which an element of interest appears, character position, ancestral path name ID, and order of branches into an element appearance information storage portion using an element name ID as a key based on the results of the analysis and registering ancestral path appearance information including at least information about the document number at which the element of interest appears, character position, element name ID, and order of branches into an ancestral path appearance information storage portion using an ancestral

path name ID as a key.

[Claim 14]

The method of claim 13, wherein the element appearance information includes at least information about the document
5 number at which the element of interest appears, character position, ancestral path name ID, order of branches, and order of empty elements, and wherein the ancestral path appearance information includes at least information about the document
10 number at which the element of interest appears, character position, element name ID, order of branches, and order of empty elements.

[Claim 15]

The method of claim 13,
wherein the registering step into the ancestral path name
15 dictionary consists of assigning a unique ancestral path name ID to each partial ancestral path name obtained by dividing each ancestral path name appearing in each structured document into more than one partial ancestral path name and registering the partial ancestral path name;
20 wherein the element appearance information includes a string of more than one ancestral path name ID instead of a single ancestral path name ID; and

wherein the ancestral path appearance information is registered in the ancestral path appearance information storage
25 portion using a string of more than one ancestral path name ID

as a key instead of a single ancestral path name ID.

[Claim 16]

The method of claim 13, further including the steps of:

grouping entries of the element appearance information

5 having common values of information items other than document
number and character position, the entries being registered in
the element appearance information storage portion using the
same element name ID as a key; and

grouping entries of the ancestral path appearance

10 information having common values of information items other than
document number and character position, the entries being
registered in the ancestral path appearance information storage
portion using the same ancestral path name ID as a key.

[Claim 17]

15 A method of searching a database for managing structured
documents by the use of a database search apparatus, the database
search apparatus having:

an element name dictionary in which an element name ID
unique to each element name appearing in each structured document

20 has been registered;

an ancestral path name dictionary in which an ancestral
path name ID unique to each ancestral path name appearing in
the structured document has been registered;

an element appearance information storage portion in which
25 element appearance information is stored using an element name

ID as a key based on results of analysis of the structured document,
the element appearance information including at least
information about a document number at which an element of
interest appears, character position, ancestral path name ID,
5 and order of branches; and

ancestral path appearance information storage portion in
which ancestral path appearance information is stored using an
ancestral path name ID as a key based on the results of the analysis
of the structured document, the ancestral path appearance
10 information including at least information about the document
number at which the element of interest appears, character
position, element name ID, and order of branches;

the method comprising the steps of:

entering a search formula;

15 converting the entered search formula into internal
condition formulas while referring to the element name dictionary
and the ancestral path name dictionary; and

finding plural search results from element appearance
information from the element appearance information storage
20 portion and from ancestral path appearance information from the
ancestral path appearance information storage portion according
to the internal condition formulas.

[Claim 18]

A database apparatus for managing structured documents,
25 the database apparatus comprising:

a database constructing apparatus having

an element name dictionary for storing an element name ID unique to each element name appearing in each structured document,

5 an ancestral path name dictionary for storing an ancestral path name ID unique to each ancestral path name appearing in the structured document,

an input document analysis portion for assigning a unique document number to the structured document and analyzing its

10 structure,

an element name registration portion for assigning a unique element name ID to each element name appearing in the structured document based on results of analysis performed by the input document analysis portion and registering the element name in

15 the element name dictionary,

an ancestral path name registration portion for assigning a unique ancestral path name ID to each ancestral path name appearing in the structured document based on the results of the analysis performed by the input document analysis portion

20 and registering the ancestral path name in the ancestral path name dictionary,

an element appearance information storage portion for storing element appearance information including at least information about document number, character position,

25 ancestral path name ID, and order of branches using an element

name ID as a key,

an ancestral path appearance information storage portion
for storing ancestral path appearance information including at
least information about document number, character position,
5 element name ID, and order of branches using an ancestral path
name ID as a key, and

an appearance information registration portion for
registering element appearance information including at least
information about the document number at which the element of
10 interest appears, character position, ancestral path name ID,
and order of branches into the element appearance information
storage portion using the element name ID of the element of
interest as a key based on the results of the analysis performed
by the input document analysis portion and registering ancestral
15 path appearance information including at least information about
the document number at which the element of interest appears,
character position, element name ID, and order of branches into
the ancestral path appearance information storage portion using
the ancestral path name ID of the element of interest as a key;
20 and

a database search apparatus having

a search condition input portion for entering a search
formula,

a search condition analysis portion for converting the
25 search formula entered by the search condition input portion

into an internal condition formula in which element name and ancestral path name are expressed by element name ID and ancestral path name ID, respectively, while referring to the element name dictionary and the ancestral path name dictionary, and

5 an appearance information acquisition portion for extracting data about plural search results complying with the internal condition formula created by the search condition analysis portion from the element appearance information stored in the element appearance information storage portion and from
10 the ancestral path appearance information stored in the ancestral path appearance information storage portion.

[Claim 19]

The database apparatus of claim 18, further including:

 an attribute name dictionary for storing attribute name
15 IDs and corresponding attribute names;

 an attribute name registration portion for assigning a unique attribute name ID to each attribute name appearing in the structured document based on results of analysis performed by the input document analysis portion and registering the
20 attribute name in the attribute name dictionary; and

 an attribute appearance information storage portion for storing attribute appearance information including at least information about document number, character position, ancestral path name ID, element name ID, and order of branches
25 using the attribute name ID as a key;

wherein the appearance information registration portion further registers attribute appearance information in the attribute appearance information storage portion using the attribute name ID as a key based on the results of the analysis performed by the input document analysis portion, the attribute appearance information including at least information about a document number at which an attribute of interest appears, character position, ancestral path name ID, element name ID, and order of branches;

wherein the search condition analysis portion further converts the search formula entered by the search condition input portion into an internal condition formula in which the attribute name is expressed by an attribute ID while referring to the attribute name dictionary; and

wherein the appearance information acquisition portion further extracts data about plural search results complying with the internal condition formula output by the search condition analysis portion from element output information stored in the element appearance information storage portion, ancestral path appearance information stored in the ancestral path appearance information storage portion, and attribute appearance information stored in the attribute appearance information storage portion.